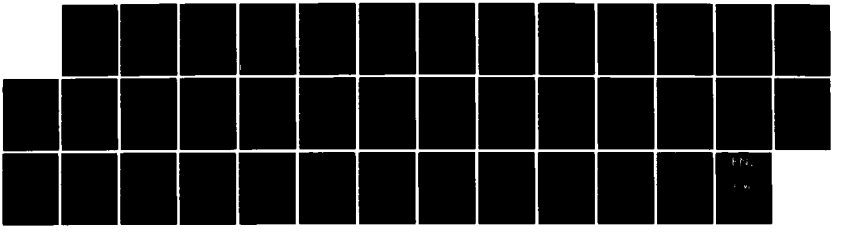
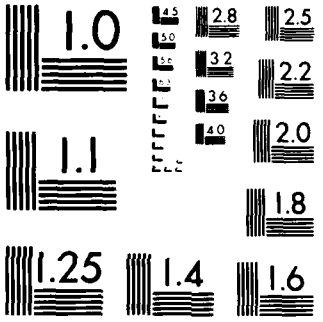
AD-A149 706 A REVIEW OF SOME ASPECTS OF ROBUST INFERENCE FOR TIME SERIES(U) WASHINGTON UNIV SEATTLE DEPT OF STATISTICS R D MARTIN SEP 84 TR-53 N00014-84-C-0169 1/1

UNCLASSIFIED F/G 12/1 NL





(12)

# A REVIEW OF SOME ASPECTS OF ROBUST INFERENCE FOR TIME SERIES

by

R. Douglas Martin

TECHNICAL REPORT No. 53
September 1984

Department of Statistics
University of Washington
Seattle, Washington 98195

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER: 53 | 2. GOVT ACCESSION NO.: AD-A149706 | 3. RECIPIENT'S CATALOG NUMBER: |
| 4. TITLE (and Subtitle): A review of some aspects of robust inference for time series | | 5. TYPE OF REPORT & PERIOD COVERED: TR 1/1/82 - 11/30/83 |
| | | 6. PERFORMING ORG. REPORT NUMBER: |
| 7. AUTHOR(s): R. Douglas Martin | | 8. CONTRACT OR GRANT NUMBER(s): N00014-84-C-0169 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, GN-22<br>University of Washington<br>Seattle, WA 98195 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-608-002 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: ONR Code N63374<br>1107 N.E. 45th Street<br>Seattle, WA 98105 | | 12. REPORT DATE: September 1984<br>13. NUMBER OF PAGES: 35 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report): Unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: |

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robustness, time series

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper briefly surveys some aspects of robust inference for time series, and gives an indication of the current state of knowledge in other problem areas. Basic notions of robustness are stated, and technical difficulties associated with the time series case are mentioned. Some models for time series with outliers are given. Least-squares procedures lack robustness for such models and robust alternatives are described. Issues of adaptivity vs. robustness are briefly mentioned. Robustness problems involving (continued)

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

dependency are briefly mentioned. Algorithms for robust data smoother-cleaners are briefly described, along with an application to radar glint noise.

# A REVIEW OF SOME ASPECTS OF ROBUST INFERENCE FOR TIME SERIES

*R. Douglas Martin**

Department of Statistics
University of Washington
Seattle, Washington 98195

*ABSTRACT*

This paper briefly surveys some aspects of robust inference for time series, and gives an indication of the current state of knowledge in other problem areas. Basic notions of robustness are stated, and technical difficulties associated with the time series case are mentioned. Some models for time series with outliers are given. Least-squares procedures lack robustness for such models and robust alternatives are described. Issues of adaptivity versus robustness are briefly mentioned. Robustness problems involving dependency are discussed. Algorithms for robust data smoother-cleaners are briefly described, along with an application to radar glint noise.

Appeared in *Statistical Signal Processing* (E. Wegman and James G. Smith, eds.), Marcel-Dekker (1984), pp 19-39.

*This research was supported by the Office of Naval Research under Contract N00014-82-K-0062.

# A REVIEW OF SOME ASPECTS OF ROBUST INFERENCE FOR TIME SERIES

*R. Douglas Martin**

Department of Statistics
University of Washington
Seattle, Washington 98195

## 1. INTRODUCTION

The body of theoretical work on time series utilizes primarily one of two mathematically convenient fictions, namely either (i) a second-order description, or (ii) a Gaussian assumption, in which a case second-order description is a complete description. The second-order formulation is at the base of many important concepts and structures in time series, including Wold's decomposition, the spectral representation, and prediction theory. In all of these one has the convenience of utilizing Hilbert space methods (for details see the appropriate sections of the recent book by Grenander, 1981). On the other hand the Gaussian assumption allows one to utilize the parametric method of maximum likelihood for time series models, early work in this area being due to Whittle (1953, 1962). *The* nonparametric method for time series consists of estimating the *spectrum*, a second-order description in the frequency domain, by a variety of methods based on the periodogram.

Unfortunately, many time series encountered in practice are quite decidedly non-Gaussian, as many practitioners know, and, correspondingly, second-order descriptions are far from adequate. Series often contain anomalies of

Appeared in *Statistical Signal Processing* (E. Wegman and James G. Smith, eds.), Marcel-Dekker (1984), pp 19-39.

*This research was supported by the Office of Naval Research under Contract N00014-82-K-0062.

numerous kinds, including local bumps or bursts, shifts in level, nonstationarities of various kinds, and isolated outliers. Least-squares and other Gaussian maximum-likelihood procedures are quite non-robust toward such phenomena. Here we shall be primarily concerned with methods which are geared to deal well with a not-too-large fraction of local bumps or bursts and isolated outliers.

It cannot be stressed too strongly that: (i) second-order descriptions are woefully inadequate for representing such phenomena, and (ii) a Gaussian marginal distribution for a series hardly insures that potent versions of such phenomena do not exist. For a striking and graphic portrayal of these two facts see the example displayed in Figures 4 through 11 of Martin and Thomson (1982). The essence of the example in these figures is that a time series often has a moderate to large amount of low frequency energy, with corresponding sample paths having broad peaks and valleys, so that outliers and bumps can be modest to small on the scale of the process (e.g., as measured by the range of the data), while being quite large on a local scale and clearly visible to the eye.

This last observation leads us to give the following loose definition of an *outlier* in a time series. An outlier $y_t$ is a data value which lies well outside of the central mass (say 95% of the mass) of conditional density $f(y_t \mid Y^{t-1})$ where the conditioning variables $Y^{t-1}$ consist of all the past observations $Y^{t-1} = (y_1, \ldots, y_{t-1})$. This density is often called the *observation prediction density*. Since we seldom get our hands on such a conditional density, it is convenient and natural to cast the definition somewhat differently. Let $\hat{y}_t^{t-1}$ denote a "good" predictor of the $y_t$ given the past $Y^{t-1}$. In particular $\hat{y}_t^{t-1}$ should have the kind of resistance/robustness properties discussed in the next section, so that this predictor is not unduly affected by outliers in $Y^{t-1}$ (such a predictor appears in Section 8.) Then $y_t$ is an outlier if the prediction residual $r_t = y_t - \hat{y}_t^{t-1}$ has magnitude large compared with a good scale measure $s_r$ for all

of the residuals $r_t$, $t = 1, \ldots, n$. For example one might well take $s_r$ to be the suitably scaled interquartile distance of the $r_t$. These definitions can be generalized in a more or less obvious way to cover the case of a "patch" or "bump" of outliers, $y_t, \ldots, y_{t+k}$.

The above comments should make the following point clear. One cannot hope to have a good method for dealing with outliers in time series by using only an instantaneous nonlinear transformation of the data, i.e., treatment of the form $\tilde{y}_t = g(y_t)$. True, some time series will contain outliers which are large on the scale of the process, and in such cases such a procedure may prevent the worst consequences. Note, however, that $\tilde{y}_t$ will in general still be an outlier in the sense given above, for this value is specified without regard to the neighboring values $y_{t-1}$, $y_{t+1}$, etc. of the series. More sophisticated procedures are called for and these will be discussed in Sections 5, 7 and 8. Sections 2 and 3 review robustness concepts for independent observations and for time series, respectively. Some time series outlier models are mentioned in Section 4. Some robust alternatives to least-squares and Gaussian maximum-likelihood procedures are introduced in Section 5. Section 6 comments on fully adaptive estimates. Section 7 deals with some aspects of robustness toward dependency, both with and without outliers simultaneously present. Finally Section 8 briefly describes robust data smoother-cleaner algorithms, and gives an application to radar glint noise.

## 2. ROBUSTNESS CONCEPTS FOR INDEPENDENT OBSERVATIONS

The following comprise four *robustness* concepts in moderately wide use today: *(1) Resistance; (2) Efficiency Robustness; (3) Min-Max Robustness; (4) Qualitative Robustness.* These concepts have been applied mainly to situations involving only independent observations until quite recently.

*Resistance*, a term due to J. W. Tukey (1976), is in fact a term distinct from *robustness*. It is the data-oriented version of the probability based word robust. As such it is the basic primitive form of robustness which captures the essential goals of robust estimation, namely *large changes* in a *smallish fraction* of the data, e.g., gross outliers, should have only a *small effect* on the estimate. *Small changes* in *all* the data, e.g., rounding (or fine quantization), should have only a *small effect* on the estimate. As is well known, least-squares and other Gaussian maximum-likelihood procedures lack resistance, and hence resistant/robust procedures have been invented.

Of the three bonafide robustness terms, the notion of efficiency robustness (Tukey, 1960; Mosteller and Tukey, 1977) is the oldest and least mathematical concept, and hence the one most accessible to applied statisticians. Let $V_S(F)$ denote a variance standard of reference at data distribution $F$, and for the moment assume we are in one of those special situations where unbiased estimates exist. $V_S(F)$ might be the Cramer-Rao bound for either asymptotic or finite-sample cases. It would preferably be the Pitman bound in the latter case, when dealing with problems such as location and scale where the Pitman bound can by some means be evaluated (Pregibon and Tukey, 1981). Alternatively, $V_S(F)$ may be simply the variance of the best known estimate at distribution $F$. With $V_T(F)$ the variance of estimate $T$ at distribution $F$, the efficiency of $T$ at $F$ is

$$EFF(T,F) = \frac{V_S(F)}{V_T(F)} . \qquad (1)$$

An *efficiency-robust estimate* $T$ is one whose efficiency is high at the nominal distribution $F_0$ (often Gaussian), and also high at strategically chosen alternative distributions $F_1, F_2, \ldots, F_K$ (usually heavy-tailed outlier-generating distributions). Often efficiencies, $REFF(T, T_{LS}; F)$, relative to least-squares or other Gaussian maximum-likelihood estimates, are used with the variance $V_{LS}$ or $V_{CMLE}$ replacing $V_S$ in (1). For problems where bias is unavoidable, and this is the case for almost all truly realistic robustness problem formulations, one will use mean-squared errors in place of variances in (1), and also compare biases as well.

Huber (1964) introduced *min-max* robust estimates in his by-now classic paper on robust estimates of location. Here the asymptotic variance $V(T,F)$ of estimator $T$ at distribution $F$ is the loss and the statistician wishes to minimize, over a family $\tilde{\mathcal{S}}$ of estimates, the maximum of $V(T,F)$ over a family $\tilde{\mathcal{F}}$ of distributions. Huber showed that such min-max estimates exist in the class of location $M$-estimates $\hat{\mu} = T$ obtained by solving

$$\min_{\mu} \sum_{1}^{n} \rho \left[ \frac{y_i - \mu}{c \cdot \hat{s}} \right] \tag{2}$$

with $\rho$ symmetric and convex, the $y_i$ independent and identically distributed (i.i.d.), and $y_i \sim F(\cdot - \mu)$. Here $\hat{s}$ is a robust scale estimate and $c$ is a tuning constant adjusted to obtain high efficiency robustness. Equivalently $\hat{\mu}$ is a solution of

$$\sum_{1}^{n} \psi \left[ \frac{y_i - \hat{\mu}}{c \cdot \hat{s}} \right] = 0 \tag{3}$$

with *psi function* $\psi = \rho'$. We henceforth choose $\hat{s} = 1$ and absorb $c$ into the definition of $\psi$ for notational convenience. Huber's (1964) famous min-max solution is based on an $\varepsilon$-contaminated family with standard Gaussian central

distribution, and the *saddle-point pair* $(T_0,F_0)$ has $T_0 = \hat{\mu}_0$ obtained from (3) with $\psi = \psi_0$ given by

$$\psi_0(t) = \begin{cases} t & |t| < K \\ K \operatorname{sgn}(t) & |t| \geq K \end{cases} \tag{4}$$

with $K = K(\varepsilon)$ determined by the contamination fraction $\varepsilon$. Other families yield other saddle-point $\psi$-functions (see for example Huber, 1981).

*Qualitative robustness* was introduced by Hampel (1968, 1971), and this is a fundamental continuity property which is the probabilistic counterpart of Tukey's data-oriented term *resistance*. Let $Y_1, \ldots, Y_n$ be i.i.d. with values in $R^k$ and common distribution $F$, and let $T_n = T_n(Y_1, \ldots, Y_n)$ define a sequence of estimates with values in $R^p$ for sample sizes $n = 1,2, \cdots$. This sequence induces the sequence of maps

$$T_n : F \rightarrow L_{T_n}(F) \tag{5}$$

where $L_{T_n}(F)$ is the *law* of $T_n$ at $F$. Then $T_n$ is said to be qualitatively robust at $F$ (or in a neighborhood of $F$, or everywhere) if the sequence of maps (5) is equicontinuous at $F$ (or in a neighberhood of $F$, or everywhere), using the Prohorov distance on the metric spaces where $F$ and $L_{T_n}(F)$ are elements. The Prohorov metric incorporates the possibility of both gross outliers and rounding errors in $\varepsilon$-neighborhoods in a natural manner, and thus is extremely attractive for use in a robustness definition.

When $\{T_n\}$ is obtained from a functional $T = T(F)$ defined on a subset $\widetilde{\mathcal{F}_s}$ of the family of all distributions by evaluation of $T$ at the empirical distribution function (e.d.f.) $F_n$, $T_n = T(F_n)$, one set of sufficient conditions for $\{T_n\}$ to be robust at $F$ is: (i) $T_n = T(F_n)$ is a continuous function on $R^n$ for each $n = 1,2, \cdots$, and (ii) $T$ is continuous at $F$. For Huber's class of location $M$-

estimations (3) $T$ is defined implicitly by

$$\int \psi(y - T(F))dF(y) = 0. \tag{6}$$

In essence robustness is achieved by choosing $\psi$ to be bounded and monotone. (In addition, uniqueness of the solution $T_0(F)$ at $F$ is needed--see Huber, 1981.)

Of the above concepts I regard resistance and qualitative robustness as fundamental, with efficiency robustness a close companion. Qualitative robustness is a principle which should be regarded on a par with other principles of statistics such as sufficiency, unbiasedness, etc. Whenever possible a statistic should be selected to have the property of qualitative robustness, all other things being relatively equal. Thus from now on the term robust, without other qualifiers, will be taken to mean qualitatively robust.

Since some rather ridiculous estimates (such as $T \equiv c$, with $c$ a constant) are robust, one needs to combine the principle with some other measure, and efficiency robustness is a natural candidate (see Beran 1977a, 1977b, for notable efforts to obtain full efficiency and robustness simultaneously).

Min-max robustness is more or less frosting on the cake: it is nice to have, but one shouldn't lose any sleep over not obtaining it. Also one should not, as has been done in some of the recent engineering literature, take min-max robustness as the *guiding* concept, at least not without some circumspection. The main justification for concentrating on min-max robustness would be that one already has a basic continuity property in hand, but that the modulus of continuity is so bad that something like a good min-max solution would be appealing. Note, however, that one must demonstrate that the modulus of continuity is indeed bad, and this is a somewhat subjective matter.

There are two important concepts affiliated with the core ideas of robustness which are also due to Hampel. The first is the *breakdown point* (Hampel,

1968, 1971), a global (asymptotic) measure which is essentially the largest fraction of contamination which an estimator can stand without breaking down completely by virtue of being taken to the boundary of the parameter space. The second concept, the *influence curve* (Hampel, 1974), is an asymptotic infinitesimal (or local) measure which gives the effect of a vanishingly small fraction of contamination of specific value on an estimate as the sample size tends to infinity.

Influence curve considerations lead one to use psi-functions (e.g., $\psi$ in Eq. (5)) that are continuous. In the sequel we take boundedness and continuity of $\psi$ to be the essential features needed for robustness. Non-monotone $\psi$ can be used by computing one-step Newton solutions to equations like (5), starting with a near-solution obtained with a monotone $\psi$.

Both the above concepts have finite-sample versions. Tukey's sensitivity curves or stylized sensitivity curves (see Andrews et al., 1972), and Mallows' empirical influence curves (Mallows, 1976) are finite sample versions of the influence curve. Hodges (1967) introduced the precursor of the breakdown point, and recently Donoho (1982) has stressed the relative importance of finite-sample breakdown points.

*Bounded-influence regression* is an approach to regression which was stimulated by the notion that an estimator's influence curve should be bounded. This problem are has seen vigorous attention by a small group of researchers (Hampel, 1975, 1978; Mallows, 1976; Krasker and Welsch, 1982; Maronna, Bustos and Yohai, 1979). This topic deserves a brief introduction, both for its own sake, and also because the approach may be adapted for robust estimation of certain times series models. Consider the regression model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \ , \quad i = 1, \ldots, n \tag{7}$$

where the $\varepsilon_i$ are i.i.d. with common symmetric distribution $F_\varepsilon$, and $\beta^T = (\beta_1, \ldots, \beta_p)$. $M$-estimates $\hat{\beta}_M$ for regression are solutions of the estimating equation

$$\sum_{i=1}^{n} \mathbf{x}_i \psi(y_i - \mathbf{x}_i^T \hat{\beta}_M) = 0 \tag{8}$$

obtained by minimizing the regression analogue of (2). It is assumed that $\psi$ is bounded, continuous and monotonic.

First suppose that the $\mathbf{x}_i$ are known *exactly* (i.e., are observed without error) and the specification (7) with regard to the $\mathbf{x}_i^T\beta$ is *correct*. Then the only source of distributional difficulty is the $\varepsilon_i$ which may contain outliers due to $F_\varepsilon$ being heavy-tailed. In this formulation $\hat{\beta}_M$ is robust according to Hampel's asymptotic definition. There may, however, still be some finite sample problems caused by so-called *X-leverage points* (see Huber, 1981, Chapter 7; Belsley, Kuh and Welsch, 1980).

On the other hand, suppose the $\mathbf{x}_i$ are occasionally observed with large errors (say keypunch errors for example), and/or the specification (7) is incorrect in any one of a variety of ways (e.g., a mixture model for $\beta$ with $P(\beta=\beta_0) = 1-\gamma$ and $P(\beta=\beta_1) = \gamma$ with $\gamma$ small). Then M-estimates $\hat{\beta}_M$ are not at all robust. In order to obtain regression estimates which are robust against such possibilities, it is desirable to use a bounded-influence (BI) regression estimate $\hat{\beta}$ which is the solution of an equation of the form

$$\sum_{i=1}^{n} \varphi(\mathbf{x}_i, y_i - \mathbf{x}_i^T \hat{\beta}) = 0 \tag{9}$$

where $\varphi(\cdot,\cdot)$ is a bounded and continuous function on $\mathbf{R}^p \times \mathbf{R}^1$. This will guard against outliers/model uncertainty in both the *independent variables*, or *carriers* $\mathbf{x}_i$ and the residuals $\varepsilon_i$. It would be quite dangerous to rely on the $M$-

estimate $\hat{\beta}_M$ if one were not quite sure about the purity of the $\mathbf{x}_i$.

The reasons for pointing out the above features of ordinary regression M-estimates and BI regression alternatives are twofold. First of all there are certain problems in communications theory (and practice) where exact knowledge of the $\mathbf{x}_i$ is virtually assured. This is the case, for example, where $\mathbf{x}_i^T\boldsymbol{\beta}$ represents a signal of known structure, such as a constant signal (i.e., a location problem) or a sinusoidal signal with unkown amplitude (where $p = 1$), or with unknown amplitude and phase (where $p = 2$). We discuss such problems in Section 7. On the other hand, when one is fitting autoregressive (AR) or autoregressive-moving-average (ARMA) models, and one has an additive outliers (AO) model, as discussed in Section 4, the *carriers* are quite definitely contaminated and observed with error. For this situation autoregression $M$-estimates are hopelessly bad, and some form of bounded-influence regression is called for.

Among the topics which deserve mention, but are otherwise beyond the scope of this paper, I would mention: (i) *quantitative robustness* (see Huber, 1981, Chapter 1); (ii) a *decision theoretic framework for robustness* (Millar, 1981); (iii) *asymptotically shrinking* $\sqrt{n}$ *neighborhood formulations* (Bickel, 1982); (iv) *finite-sample min-max results for testing and confidence intervals* (Huber, 1981, Chapter 10); (v) *Hampel's extremal problem* (Huber, 1981, Chapter 11).

## 3. ROBUSTNESS CONCEPTS FOR TIME SERIES

Although the fundamental continuity idea behind robustness has a simple and immediate appeal, both the definition and the proofs of sufficient conditions are highly technical (even the need for the *equi*continuity part of the definition requires a little explanation). This is unfortunate because it makes all levels of detail quite inaccessible to the practitioner or engineer. *Resistance* is a much more palatable concept in this regard, but even this concept may require careful verification for complex estimates. Things get even more complicated when one tries to provide an adequate definition of qualitative robustness for time series problems.

On the other hand, it is quite important to have a solid theory as a cornerstone from which to build. If the theory is complex, as is now the case, then the theoretician has a responsibility to communicate the central concepts and results as clearly and simply as possible to potential users of proposed robust procedures.

### *Parameter Estimation*

In recognition of the need for a suitable version of qualitative robustness for time series parameter estimates, the following researchers have made contributions to the problem: Papantoni-Kazakos and Gray (1979), Cox (1981), Bustos (1981) and Boente, Fraiman and Yohai (1982).

An issue arising in the time series case is that of specifying the metric, and hence the topology, for the space of sample paths. There are a variety of ways to do this, as is reflected in the above references, and what is required is a reasonable balance so that the topology is neither too weak (in which case no estimates are robust) nor too strong (in which case all estimates are robust).

Papantoni-Kazakos and Gray (1979) work with the so-called $\overline{\rho}$ (rho-bar) metric. Their definition has a defect in the arbitrariness of the per-letter metric $\rho_0$ used to arrive at a final $\overline{\rho}$ metric. In order to deal with arbitrarily heavy-tailed processes, for example, it is necessary to choose $\rho_0$ bounded. Cox's (1981) definition circumvents this difficulty, but only applies to estimates whose functional versions (analogous to $T(F)$ in (6)) depend on only a finite-dimensional marginal distribution for the process.

The Boente, Fraiman and Yohai (1982) work, initiated by Yohai, seems to be the most attractive. A major feature of their definition is that the metric $d_\gamma^n$ they use for sample paths of length $n$ is extremely natural and transparent:

$$d_\gamma^n = \inf\left\{\gamma : \frac{\#\{i : |y_i - y'_i| \geq \gamma\}}{n} \leq \gamma\right\} \tag{10}$$

where $\#\{i : |y_i - y'_i| \geq \gamma\}$ is the number of coordinates in the two observed sample paths $y = (y_1, \ldots, y_n)$ and $y' = (y'_1, \ldots, y'_n)$ which differ by at least $\gamma$. Thus $d_\gamma^n$ is the smallest $\gamma$ such that the fraction of coordinates whose difference exceeds $\gamma$ is no greater than $\gamma$. This is a data-based distance which allows for both rounding up to an amount $\gamma$, and a fraction $\gamma$ of gross errors in a $\gamma$ neighborhood. Of course the final definition of robustness involves some additional structure, and also letting $n \to \infty$.

Consider an estimate $T_n$ obtained by solving the estimating equation of rather general form

$$\sum_{i=1}^{n} \psi_i(y'_1, \ldots, y'_n; T_n) = 0 \tag{11}$$

where $y'_1, \ldots, y'_n$ is the observed segment of a time series. The essential requirement needed to insure robustness is that the psi-functions $\psi$ be bounded and continuous. Specific examples are given in Section 5.

*Filtering and Smoothing Problems*

In filtering and smoothing problems we have as many estimates, call them $\hat{x}_t$, $t = 1, \ldots, n$, as there are observed data values $y_1, \ldots, y_n$. Thus a filter or a smoother is a mapping $S_n$ from $R^n$ to $R^n$. It is not clear exactly what constitutes an appropriate definition of qualitative robustness for problems of this type. We surely want some form of continuity for the sequence of maps $\tilde{S}_n : \mu \to \mu_{S_n}(u)$ where $\mu$ is the measure for the stationary process $y_t$ and $\mu_{S_n}(\mu)$ is the measure for $\hat{x}_1, \ldots, \hat{x}_n$. Consistency is not a possibility in filtering and smoothing problems, and evidently equicontinuity may not be as crucial here. However, this remains to be determined.

At the very least, we would require a resistance version of robustness for the $\hat{x}_t$, $t = 1, \ldots, n$. This amounts to requiring that the map $S_n$ defines a bounded and continuous functional of $\mu_n$, the measure for $y_1, \ldots, y_n$. Boundedness insures that no single $y_t$ can spoil the $\hat{x}_t$, and continuity insures that small rounding errors cannot have a large effect. Thus we would require that $S_n = S_n(\mu_n)$ be a *weakly continuous function* on the space $\tilde{\mathcal{G}}_n$ of measures $\mu_n$ for $y^T = (y_1, \ldots, y_n)$. (Compare this with Huber, 1981, Chapter 1.) Linear filters and smoothers lack resistance--appropriate bounded and continuous nonlinearity is required to achieve robust/resistant filters and smoothers. The smoother-cleaners of Section 8 have this property.

## 4. TIME SERIES MODELS FOR OUTLIERS

In some previous work I have concentrated on the robust estimation of AR and ARMA model parameters, and robust spectral density estimation, utilizing the following two distinct outlier generating models for observed time series $y_t$ (see Martin, 1981, and Martin and Thomson, 1982, and the references therein).

*The Innovations Outliers (IO) Model*

$$x_t = \mu + \sum_{l=0}^{\infty} h_l \varepsilon_{t-l} \tag{12}$$

where the $\varepsilon_t$ are i.i.d. with common distribution $F$ which is symmetric and possibly heavy-tailed, $\sum h_l^2 < \infty$ and $\mu$ is the *location* parameter for $x_t$. Then let

$$y_t = x_t \tag{12'}$$

be perfect observations of the $x_t$ process.

*The Additive Outliers (AO) Model*

$$x_t = \mu + \sum_{l=0}^{\infty} h_l \varepsilon_{t-l} \tag{13}$$

with $\varepsilon_t$ i.i.d. Gaussian, $\sum h_l^2 < \infty$ and

$$y_t = x_t + v_t \tag{14}$$

where $P(v_t=0) = 1-\gamma$ with $\gamma$ small. The AR and ARMA models are special cases of the general linear processes (12) and (13).

For the AR case the IO model corresponds roughly to a finite parameter linear regression model with heavy-tailed error distribution. However, some quirks of the model exist, and will be mentioned in the next section. The $v_t$ in the AO model represent outliers, either in patches or in isolation, and in the AR case we have the analogue of a linear regression model with Gaussian residuals, but with errors in the variables (EV).

The AO model is a special case of a more general kind of $x_t$ perturbation model

$$y_t = (1-z_t)x_t + z_t w_t \tag{15}$$

with $z_t$ a binary series with $P(z_t=1) = \gamma$ (see Yohai and Bustos, 1982). We shall also refer to this as an AO model, even though the term replacement model might equally well be used.

*ARCH Autoregressions*

Recently we have also been studying the properties of the following type of ARCH autoregressions and associated parameter estimation problems (Nemec and Martin, 1983). Let

$$y_t = \gamma + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \tag{16}$$

with $\varepsilon_t$ an ARCH process as defined by Engle (1981):

$$\varepsilon_t | \underline{\varepsilon}^{t-1} \sim N(0, h(\underline{\varepsilon}^{t-1})) \tag{17}$$

where $\underline{\varepsilon}^{t-1}$ is the past history of the $\varepsilon_t$. The intercept $\gamma$ accounts for a non-zero mean for $y_t$. The $\varepsilon_t$ are uncorrelated, but *not* independent. The functions $h$ which we have concentrated on are of the same form which Engle (1981)

emphasizes in the regression context:

$$h(\mathcal{E}^{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 \quad . \tag{18}$$

The parameters $\alpha_i$ must satisfy certain minimal constraints to insure wide-sense stationarity, and more severe constraints to insure existence of higher order moments (see Engle, 1981). The usual Gaussian autoregression is a special case of (18) obtained by $\alpha_1 = \cdots = \alpha_p = 0$, and $\alpha_0 = \sigma^2$.

The marginal density for $\varepsilon_t$ is more or less heavy-tailed, depending on the values of the $\alpha_i$. This statement may be inferred by checking that certain higher order moments do not exist, depending on the values of the $\alpha_i$, and by empirical checks based on the (easily) simulated ARCH type $\varepsilon_t$. None-the-less, an open problem concerning the $\varepsilon_t$ process itself is that of determining an analytic form for the stationary distribution of the $\varepsilon_t$, even in the simplest case where

$$h(\varepsilon^{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \quad .$$

ARCH autoregressions are potentially much more useful than IO autoregressions mainly because their sample paths seem more realistic representations of many time series sample paths arising in practice.

*Regression with Non-Gaussian AR Residuals*

In Section 7 we discuss robust point estimation of $\boldsymbol{\beta}$ in the following model:

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta} + u_t \tag{19}$$

with the very special assumptions that the $x_t$ are known exactly, and

$$u_t = \varphi_1 u_{t-1} + \cdots + \varphi_p u_{t-p} + \varepsilon_t \tag{19'}$$

where $\varepsilon_t$ is a possibly heavy-tailed outlier producing mechanism. The $\varepsilon_t$ could be i.i.d., or an ARCH process. This setup includes the special case of estimating location with non-Gaussian AR errors. Except for the location case where some work has been done (Portnoy, 1977; Wegman and Carroll, 1977), this problem has not been studied at all in the previous literature.

## 5. LEAST-SQUARES AND ROBUST ESTIMATES OF AUTOREGRESSIONS

Let's focus solely on the autoregression versions of IO and AO models, and the AR ARCH models described in the previous section. Discussion of moving average models is omitted here for the sake of brevity. A perfectly observed Gaussian autoregression is regarded as the nominal model, with IO, AO and AR ARCH models particular types of non-Gaussian deviations from this nominal model.

Consider the $p$th-order autoregression version of the regression $M$-estimate (8) for a $\mu = 0$ version of (12) and (13):

$$\sum_{t=p+1}^{n} \mathbf{z}_t \psi(y_t - \mathbf{z}_t^T \hat{\boldsymbol{\varphi}}_M) = 0 \tag{20}$$

where $\mathbf{z}_t^T = (y_{t-1}, \ldots, y_{t-p})$. This includes the least-squares estimate $\hat{\boldsymbol{\varphi}}_{LS}$ as a special case. Now $\hat{\boldsymbol{\varphi}}_{LS}$ has a rather notable property at finite variance IO models: its asymptotic covariance matrix depends only upon $\boldsymbol{\varphi}$, and not upon the distribution of the $\varepsilon_t$ (Whittle, 1962; Martin, 1982a). This was cited as a robustness property by Whittle.

However, several points are in order. First of all, unlike $\hat{\boldsymbol{\varphi}}_M$, $\hat{\boldsymbol{\varphi}}_{LS}$ lacks efficiency robustness at IO models (Martin, 1982). Secondly $\hat{\boldsymbol{\varphi}}_{LS}$ is disastrously non-robust toward AR ARCH models (Nemec and Martin, 1983). We conjecture that $\hat{\boldsymbol{\varphi}}_M$ is robust toward AR ARCH models, but this remains to be established. More importantly, neither $\hat{\boldsymbol{\varphi}}_{LS}$ or $\hat{\boldsymbol{\varphi}}_M$ are robust toward AO models of either the specific type (14) or the general type (15); both type of estimates suffer from severe biases as well as inflated variances (Denby and Martin, 1979).

Since AO models are included in arbitrarily small Prohorov neighborhoods of a Gaussian autoregression (see, for example, Cox, 1981) both $\hat{\boldsymbol{\varphi}}_{LS}$ *and* $\hat{\boldsymbol{\varphi}}_M$ lack qualitative robustness! Following the comments made in conjunction with (11),

we require estimating equations whose summands are bounded and continuous functions of the data, and this is not the case with the $M$-estimate defined by (20). The point is that AO models give rise to errors in the $\mathbf{z}_t$ which can have quite potent effects.

Three classes of robust estimates have been proposed for this setup: *(i) Bounded-Influence Autoregression (BIFAR); (ii) RA-Estimates; (iii) Robust Data Cleaning followed by Least-Squares.* The first class utilizes *bounded-influence regression* type estimates, or generalized $M$-estimates ($GM$-estimates) applied to autoregressions. The two main variants are the Hampel-Krasker-Welsch version and the Mallows version (see Martin, 1981, and the references therein).

The second class of estimates, due to Yohai and Bustos (1982), are obtained as follows. First, one computes *robust* covariances $\tilde{\gamma}_k = \tilde{\gamma}_k(\varphi)$ of lag-$k$ residuals:

$$\tilde{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} \tilde{\psi}(r_t, r_{t+k}) \tag{21}$$

where $r_t = r_t(\hat{\varphi}) = y_t - (\hat{\varphi}_1 y_{t-1} + \cdots + \hat{\varphi}_p y_{t-p})$ are the residuals. Then the $\tilde{\gamma}_k$ are substituted for the conventional covariance estimates $\hat{\gamma}_k$, obtained when $\tilde{\psi}(r_t, r_{t+k}) = r_t \cdot r_{t+k}$, in the usual least-squares equations expressed in terms of $\hat{\gamma}_k$ (see Yohai and Bustos, 1982, for details).

Robustness is achieved by choosing $\tilde{\psi}$ to be a bounded and continuous function on $R^2$. One choice for $\tilde{\psi}$ is $\psi(u,v) = \psi(u)\psi(v)$ for some bounded, continuous $\psi$ function on $R^1$. The essential idea is that the estimates yield zero values for robust lag-$k$ correlation estimates of the residuals, for $k = 1, \ldots, p$ incorporated in a manner which results in high efficiency. Hence the name *RA-estimates* stands for (robust) residual-autocorrelation-based estimates.

The third class of estimates is obtained by iterative application of a *robust smoother-cleaner* to remove outliers, followed by application of the usual least-

squares estimate (Kleiner, Martin and Thomson, 1979; Martin, Samarov and Vandaele, 1982; Martin and Thomson, 1982). The smoother-cleaner has the property that at a gross-outlier position (in the sense described in Section 1), the outlier is replaced by an interpolate based on all the other cleaned data. An algorithm for smoother-cleaners is given in Section 8. Robustness is obtained for this method by virtue of the smoother-cleaner being a bounded and continuous function of the data.

All three of the above classes of estimates may be modified to cover the case of nominally Gaussian ARMA models with varying degrees of elegance, and success yet to be fully determined.

A careful comparative study of the three approaches is not yet available. Yohai and Bustos (1982) should have good comparative results on classes (1) and (2) for AR(1) and MA(1) models in the very near future. Both BIFAR and RA estimates are consistent and highly efficient at the nominal Gaussian AR model (Fisher consistency), while being robust for well chosen psi-functions. They are typically asymptotically normal as well, and have small biases at AO models (one might well call this latter feature *bias robustness).* I believe that the RA-estimates will be generally preferred to BIFAR estimates for at least two good reasons aside from their efficiency and bias robustness. Assuming the latter are on at least a roughly even par with BIFAR estimates, the RA-estimates are (i) quite natural for time series models, and can be applied in principle to models of considerable complexity, and (ii) they can be designed with just one efficiency tuning constant whose values are relatively easy to determine (compare this with the difficulty involved in choosing tuning constants for BIFAR estimates implied by Peters, Samarov and Welsch's (1982) discussion in the general regression context).

The method of robust data-cleaning, followed by least squares in an iterative manner, is a quite natural and attractive one. Note, however that it requires the use of a BIFAR or RA-estimate to provide a reasonably good starting point for iteration, as the overall procedure is highly nonlinear. It is even some kind of approximation to a non-Gaussian M.L.E. if an appropriate filter-smoother is used (Martin, 1981), and it fits in nicely with a robust prewhitening approach to spectral density estimation (Kleiner, Martin and Thomson, 1979; Martin and Thomson, 1982). The method has a drawback whose importance is somewhat debatable, namely the method is not Fisher consistent. This is certainly quite objectionable from a theoretical point of view, and there unfortunately seems to be no easy way to get around the problem other than through some form of adaption. This we intend to pursue in the near future. On the other hand certain calculations show that the asymptotic bias at the nominal Gaussian model will be so small as to have little practical consequence (Martin and Thomson, 1982, Section 6).

## 6. FULL ADAPTION VERSUS ROBUSTNESS

During the course of the workshop for which this talk was prepared, the following extemporaneous remarks were made.

Some attention was given by several speakers to density estimation and score function approximation, where the *(efficient) score function* is $\Psi = -f'/f$, $f$ being a density for presumably i.i.d. data. Such attention is presumably motivated by a desire to use blatantly adaptive methods. This prompted recollection of Stone's (1975) Monte Carlo results presented at the end of his asymptotic treatment of adaptive, asymptotically efficient, location estimates $\hat{\mu}$. These estimates are obtained by solving

$$\sum_{1}^{n} \hat{\Psi}_n \left[ \frac{y_i - \hat{\mu}}{\hat{s}} \right] = 0 \tag{22}$$

where $\hat{\Psi}_n$ is an estimate of $\Psi$ and $\hat{s}$ is a robust scale estimate. Stone used $\hat{\Psi}_n(r) = [-\hat{f}'_n(r)/\hat{f}_n(r)] \cdot d_n(r)$ where $\hat{f}_n$, $\hat{f}'_n$ are kernel density estimates using a Gaussian density type kernel, and $d_n(r)$ truncates $[-\hat{f}''_n(r)/\hat{f}_n(r)]$ to zero outside a symmetric interval $[-a_n, a_n]$ with $a_n \to \infty$ as $n \to \infty$.

A question frequently raised about such fully adaptive estimates is, "How large must $n$ be in order for the asymptotics to set in?" Somewhat surprisingly, $n$ needn't be so large, as Stone's Monte Carlo for sample size $n = 40$ showed. His results give $EFF(\hat{\mu}, f) \geq 0.89$ for $f$ ranging over the Gaussian, Laplace, Contaminated Normal (contamination fraction = 0.1, contamination variance = 9) and Cauchy distributions.

While Stone's Monte Carlo results are quite encouraging, his results need to be contrasted with the fact that: (i) comparable results are achieved with a robust location $M$-estimate of the type (3) using a good $\hat{s}$, an appropriate value for $c$, and a good redescending psi-function $\psi$--for example Tukey's bisquare

psi-function (see Mosteller and Tukey, 1977); and (ii) such an $M$-estimate is computationally *much* simpler than the fully adaptive estimate (22).

It is doubtful that there are many applications where going the additional 10% or so, from around 90% to full efficiency, is worth the computational effort and complexity of the fully adaptive estimate. A counter argument is that if staying as close as possible to full efficiency is really cheap, then why not? Of course we should really check to determine at what (small) sample size full adaption becomes untenable.

## 7. ROBUSTNESS AND DEPENDENCY

In this section we wish to make two main points. The first is that relatively small amounts of serial correlation can seriously affect the level (or false alarm) of a test, or equivalently the error rate of a confidence interval. This is true even in the completely Gaussian case, where it is a surprisingly unadvertised fact that tests and confidence intervals are very non-robust toward dependency. Here we use the word robust very loosely and intuitively--the definitions of qualitative robustness for time series given in Section 3 may need to be modified for this kind of problem.

The second point is made in connection with the *very special* model assumptions made in equations (19)-(19'). Namely, ordinary location *M*-estimates are not adequate for estimation of location with non-Gaussian autoregressive errors, unless the dependency is quite weak. They can be quite inefficient compared with *proper M-estimates*, i.e., true M.L.E. type estimates for the actual model. Similar comments apply to problems of linear regression with non-Gaussian autoregressive errors.

*The Student's t Confidence Interval with Dependency*

Consider the usual Student's $t$ 95% confidence interval which has error rate of 5%: $CI = (\overline{y} - t_{.025,n-1} S/\sqrt{n},\ \overline{y} + t_{.025,n-1} S/\sqrt{n})$, where $\overline{y}$ is the sample mean of $y_1, y_2, \ldots, y_n$, and $S^2$ is the usual sample variance estimate. Suppose that in fact the $y_t$ are given by the special case of (19)-(19') where $x_t^T \beta = \mu$, a location parameter, and that $u_t$ in (19') is a zero mean Gaussian AR(1) process with transition parameter $\varphi$. If in fact $\varphi = 0$, then CI has the stated error rate of 0.05. However when $\varphi \neq 0$ and the sample size is large, the results are as follows:

| $\varphi$ | Error Rate |
|---|---|
| 0.25 | 0.13 |
| 0.5 | 0.27 |
| 0.7 | 0.42 |
| 0.9 | 0.66 |

The results are dramatic. For $\varphi = 0.25$ the error rate has more than doubled, and things get rapidly worse with increasing $\varphi$. The problem is that as $n \to \infty$

$$S^2 \to \text{VAR}\, y_1 = \frac{\sigma_\varepsilon^2}{1-\varphi^2} \neq \text{VAR}_\infty \sqrt{n}\, \bar{y} = \frac{\sigma_\varepsilon^2}{(1-\varphi)^2} = S_u(0) \qquad (23)$$

where $\text{VAR}_\infty$ denotes the asymptotic variance, and $S_u(f)$ is the spectral density for the error process $u_t$. It should be noted that the right hand equalities hold quite generally; we needn't restrict ourselves to AR or even ARMA processes (Grenander, 1981). What we need to do to studentize $\bar{y}$ with dependency present is an estimate of $S_u(0)$, the spectral density of the error process at the origin. The same is true with regard to setting the threshold for tests.

Heidelberger and Welch (1980) have studied nonparametric methods for doing this. The author and a student have checked the behavior of autoregressive type estimates of $S_u(0)$ with Akaike's (1977) order selection rule AIC, in a casual way via Monte Carlo. This also seems to work with the proviso that jackknifing must be done to remove the $0(n^{-1})$ bias in the autoregressive coefficient estimates if the sample size is not large enough relative to the amount of correlation (this remains to be determined with care, but for an AR(1) process, $\varphi = 0.8$ and $n = 50$ definitely requires such bias removal).

*Robust Estimation of Location*

P. Huber's (1964) $M$-estimates $\hat{\mu}_{OM}$ of location, obtained by solving (3), were introduced in the context of independent and identically distributed observations $y_t$. The new subscript notation "$OM$" stands for *ordinary* location $M$-estimate, for reasons which will become obvious shortly. The behavior of $\hat{\mu}_{OM}$ when the $y_t$ are both dependent and non-Gaussian has received relatively little attention. However, some relatively recent work includes that of Portnoy (1977) and Wegman and Carroll (1977). The main conclusions of Portnoy's work are: (i) if the $y_t$ have only weak correlation structure then $\hat{u}_{OM}$ has high absolute efficiency for heavy-tailed distributions associated with moving-average type errors; (ii) weak dependency and heavy-tailedness seems to motivate the use of redescending psi-function.

Unfortunately, ordinary location $M$-estimates cannot compete with *proper* location $M$-estimates with non-Gaussian ARMA model errors when the correlation structure is moderate to strong. By *proper* $M$-estimate we mean true maximum-likelihood type estimates appropriate for the model. These are obtained as follows.

Let $y_t$ by given by the location model special case of (19)

$$y_t = \mu + u_t \tag{24}$$

where the $u_t$ are now an ARMA $(p,q)$ generalization of (19') process

$$u_t + \varphi_1 u_{t-1} + \cdots + \psi_p u_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \ . \tag{24'}$$

Heavy-tailed $F$'s give rise to outliers in the $\varepsilon_t$, and hence in the $u_t$ and $y_t$. This model may be written in the equivalent form

$$y_t + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} = \gamma + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{25}$$

where the expression for the intercept $\gamma$ is

$$\gamma = \mu(1 + \Sigma\varphi_i) \quad . \tag{26}$$

Let $\alpha = (\gamma, \varphi, \theta)$ denote the true parameter vector for (24)-(24') or (25)-(26), and let $\alpha'$ denote an arbitrary value in the region where the process $y_t$ is stationary and invertible. For a given $\alpha'$ one can generate residuals $r_t(\alpha')$ from the recursion, using appropriate initial conditions, in the usual way (see for example, Box and Jenkins, 1976). An $M$-estimate $\hat{\alpha}$ of $\alpha$ is a solution of the minimization problem

$$\min_{\alpha'} \sum_{t=1}^{n} \rho\left[\frac{r_t(\alpha')}{c \cdot \hat{s}}\right] \quad . \tag{27}$$

For $\rho(t) = -\log f(t)$, this yields a conditional maximum likelihood estimate (conditioned on $y_1, \ldots, y_p$ and the initial conditions for the $\varepsilon_t$), which is asymptotically efficient. Consistency and asymptotic normality of "one-step" $M$-estimates are established in Lee and Martin (1983).

Now given the $M$-estimates $\hat{\alpha} = (\hat{\gamma}, \hat{\varphi}, \hat{\theta})$, the relation (26) leads to the *proper* location $M$-estimate

$$\hat{\mu} = \frac{\hat{\gamma}}{1 + \Sigma\hat{\varphi}_i} \quad . \tag{28}$$

In the special case where $\rho(t) = -\log f(t)$ this yields the conditional M.L.E. of $\mu$. The above estimate is the one which is really the appropriate $M$-estimate of $\mu$ for the model (24)-(24').

Detailed comparisons of the asymptotic and finite sample behaviors of $\hat{\mu}_{OM}$ and $\hat{\mu}_M$ are given for AR(1) and MA(1) models by Lee and Martin (1983). It is shown that the efficiency of $\hat{\mu}_{OM}$ can be quite small relative to that of $\hat{\mu}$.

*Robust Estimation of Signal Parameters*

The regression model (19) contains as special cases some of the classical models of communication theory, where one is estimating signal parameters. For example, estimation of signal amplitude deals with the case $\mathbf{x}_t^T = \beta \cos 2\pi f_0 t$, while estimation of signal amplitude and phase is based on the case where $\mathbf{x}_t^T = \beta_1 \cos 2\pi f_0 t + \beta_2 \sin 2\pi f_0 t$. For these models it turns out that the ordinary least-squares estimates are asymptotically efficient when the $\varepsilon_t$ in (19') are Gaussian, and even under much more general assumptions for Gaussian $u_t$ (Grenander and Rosenblatt, 1957; Grenander, 1981).

However, when the $\varepsilon_t$ are non-Gaussian and heavy-tailed, the situation is much the same as in the location problem just discussed. An alternative to least squares is required, but *ordinary* M-estimates lack efficiency robustness. One requires a *proper* M-estimate geared to the model (19)-(19'), and such estimates are unfortunately a bit more complicated than in the simple case of estimating location. One possibility for computing proper M-estimates for regression models with non-Gaussian AR errors is via a straightforward robustification of Durbin's (1960) two-stage least-squares procedure. Details may be found in Martin (1982b).

## 8. ROBUST DATA SMOOTHER CLEANERS

As was mentioned in Section 5, so-called *smoother-cleaners* form a building block for robust parameter estimation. They also form a basis for robust spectral estimation via a robust prewhitening approach. Since details are provided in the references cited in Section 5, only the briefest of descriptions and an example are provided here.

Consider the AO model (14), with $x_t$ and AR(p) process having a state-variable representation $\mathbf{X}_t = \Phi \mathbf{X}_{t-1} + \mathbf{U}_t$, with $x_t = (\mathbf{X}_t)_1$ being the first component of the $p$-vector $\mathbf{X}_t$, and similarly $\varepsilon_t = (\mathbf{U}_t)_1$. In the first pass the data $y_t$ is processed in forward time with the *filter-cleaner* algorithm

$$\hat{\mathbf{X}}_t = \Phi \hat{\mathbf{X}}_{t-1} + \mathbf{m}_t s_t \psi \left[ \frac{y_t - \hat{y}_t^{t-1}}{s_t} \right] \tag{29}$$

where

$$\hat{y}_t^{t-1} = (\Phi \hat{\mathbf{X}}_{t-1})_1 \tag{30}$$

is a robust one-step-ahead predictor, as was mentioned in Section 1; here $\psi$ is bounded and continuous, and the "gain" $\mathbf{m}_t$ and the time-varying scale $s_t$ are computed from auxiliary recursions. In essence (29)-(30) is a robustified Kalman filter with data-dependent gain and scale sequences.

The smoother-cleaner output is then obtained by the reverse-time pass

$$\hat{\mathbf{X}}_t^n = \hat{\mathbf{X}}_t + A_t(\hat{\mathbf{X}}_{t+1}^n - \Phi \hat{\mathbf{X}}_t) \ , \quad t = n-1, n-2, \ldots, 1 \tag{31}$$

with initial condition $\hat{\mathbf{X}}_n^n = \hat{\mathbf{X}}_n$. Here the $\hat{\mathbf{X}}_t$ come from (29), and the $A_t$ are computed from quantities appearing in the auxiliary recursions for (29). This algorithm is a robustified form of the optimal linear smoothing algorithm due to Meditch (1969).

As an example of the efficacy obtainable through use of the smoother-cleaner (29)-(30), consider the *glint noise* sample path in Figure 1. This highly spikey non-Gaussian data is obtained from radar measurements of position of an aircraft target. The composite, reverberation-like nature of the radar return is the cause of the glint spikes, which result in an unnecessarily high observation-noise variance at the input of a target tracking loop. These spikes can be nicely eliminated, and the observation noise level thereby tremendously reduced, by use of a smoother-cleaner, as shown in Figure 2, where a 3rd-order autoregressive approximation for the data was used. For details concerning the application of smoother-cleaners to glint noise data, see Section VII of Martin and Thomson (1982).

## REFERENCES

Akaike, H. (1977), "On entropy maximization principle," *Proceedings of the Symposium on Applications of Statistics*, Wright State Univ., Dayton, OH.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972), *Robust Estimates of Location–Survey and Advances*, Princeton University Press, Princeton, NJ.

Belsley, D. D., Kuh, E. and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York.

Beran, R. (1977a), "Robust location estimates," *Annals Statist.*, *5*, 431-444.

Beran, R. (1977b), "Minimum Hellinger distance estimates for parametric models," *Annals Statist.*, *5*, 445-463.

Bickel, P. J. (1982), "Robust regression based on infinitesimal neighborhoods," Tech. Rep. No. 16, Dept. of Statistics, Univ. of California, Berkeley.

Boente, G., Fraiman, R. and Yohai, V. (1982), "Qualitative robustness for general stochastic processes," Tech. Rep. No. 26, Dept. of Statistics, Univ. of Washington, Seattle.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Bustos, O. (1980), "Qualitative robustness for general processes," unpublished manuscript, Inst. de Matematica Pura y Aplicada.

Cox, D. (1981), "Metrics on stochastic processes and qualitative robustness," Tech. Rep. 3, Dept. of Statistics, Univ. of Washington, Seattle.

Denby, L. and Martin, R. D. (1979), "Robust estimation of the first order autoregressive parameter," *Jour. Amer. Stat. Assoc.*, *74*, No. 365, 140-146.

Donoho, D. (1982), "Breakdown properties of multivariate location estimators," Ph. D. Qualifying Paper, Statistics Dept., Harvard Univ., Cambridge, MA.

Durbin, J. (1960), "Estimation of parameters in time series regression models," *Jour. Royal Stat. Soc.*, *B*, *22*, 139-153.

Engle, R. F. (1981), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," discussion paper (second revision), Dept. of Economics, Univ. of California, San Diego.

Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.

Grenander, U. (1981), *Abstract Inference*, Wiley, New York.

Hampel, F. R. (1968), "Contributions to the theory of robust estimation," Ph. D.

Dissertation, Dept. of Statistics, Univ. of California, Berkeley.

Hampel, F. R. (1971), "A general qualitative definition of robustness," *Annals Math. Stat.*, *42*, 1887-1895.

Hampel, F. R. (1974), "The influence curve and its role in robust estimation," *Jour. Amer. Stat. Assoc.*, *69*, 383-393.

Hampel, F. R. (1975), "Beyond location parameters: robust concepts and methods," *Proceedings of I.S.I. Meeting, 40th Session*, Warsaw.

Hampel, F. R. (1978), "Optimally bounding the gross-error-sensitivity and the influence of position in factor space," *1978 Proceedings of the A.S.A. Statistical Computing Section*, Amer. Stat. Assoc., Washington, DC.

Heidelberger, P. and Welch, P. D. (1980), "A spectral method for simulation confidence interval generation and run control length," Research Report RC 8264 (#35526), IBM Thomas J. Watson Research Center, Yorktown Heights, NY.

Hodges, J.L. (1967), "Efficiency in normal samples and tolerance of extreme values for some estimates of location," *Proceedings Fifth Berkeley Symposium on Math. Statistics and Prob.*, Vol. 1, 163-186.

Huber, P. (1964), "Robust estimation of a location parameter," *Annals Math. Stat.*, *35*, 73-101.

Huber, P. (1981), *Robust Statistics*, Wiley, New York.

Kleiner, B., Martin, R. D. and Thomson, D. J. (1979), "Robust estimation of power spectra" (with discussion), *Jour. Royal Stat. Soc., B, 41*, No. 3, 313-351.

Krasker, W. S. and Welsch, R. E. (1982), "Efficient bounded-influence regression estimation," *Jour. Amer. Stat. Assoc., 77*, No. 379, 595-604.

Lee, C. H. and Martin, R. D. (1982), "M-estimates for ARMA processes," Tech. Rep. No. 23, Dept. of Statistics, Univ. of Washington, Seattle.

Lee, C. H. and Martin, R. D. (1983), "M-estimates of location for dependent data," Tech. Rep. No. 29, Dept. of Statistics, Univ. of Washington, Seattle.

Mallows, C. L. (1976), "On some topics in robustness," Bell Labs. Tech. Memo, Murray Hill, New Jersey. (Talks given at NBER Workshop on Robust Regression, Cambridge, MA, May 1973, and at ASA-IMS Regional Meeting, Rochester, NY, May 21-23, 1975.)

Maronna, R. A., Bustos, O. and Yohai, V. (1979), "Bias and efficiency robustness of general M-estimates with random carriers," *Smoothing Techniques for Curve Estimation*, (Proceedings, Heidelberg, 1979), edited by T. Gasser and M. Rosenblatt, Springer-Verlag, New York.

Martin, R. D. (1981), "Robust methods for time series," in *Applied Time Series II*, edited by D. F. Findley, Academic Press, New York.

Martin, R. D. (1982a), "The Cramer-Rao bound and robust M-estimates for

autoregressions," *Biometrika, 69.*

Martin, R. D. (1982b), "Robust estimation of signal parameters with dependent data," *Proceedings 21st IEEE Conference on Decision and Control,* 433-436.

Martin, R. D., Samarov, A. and Vandaele, W. (1982), "Robust methods for ARIMA models," to appear in proceedings of ASA-CENSUS-NBER Conference on Applied Time Series Analysis of Economic Time Series, Oct. 13-15, 1981.

Martin, R. D. and Thomson, D. J., (1982), "Robust resistant spectrum estimation," *IEEE Proceedings,* Vol. 70, No. 9, 1097-1115.

Meditch, J. S. (1969), *Stochastic Optimal Linear Estimation and Control,* McGraw-Hill, New York.

Millar, P. W. (1981), "Robust estimation by minimum distance methods," *Z. Wahrschein., 55,* 73-89.

Mosteller, F. and Tukey, J. W. (1977), *Data Analysis and Regression,* Addison-Wesley, Reading, MA.

Nemec, A. and Martin, R. D. (1983), "Some properties of some ARCH type autoregressions and least squares estimates," in preparation.

Papantoni-Kazakos, P. and Gray, R. M. (1979), "Robustness of estimators on stationary observations," *Annals Probability, 7,* No. 6, 989-1002.

Peters, S. C., Samarov, A. and Welsch, R. E. (1982), "Computational procedures for bounded-influence regression," *Proceedings of the 1982 INTERFACE Meeting,* Troy, NY.

Portnoy, S. L. (1977), "Robust estimation in dependent situations," *Annals Statist., 4,* 1127-1137.

Pregibon, D. and Tukey, J. W. (1981), "Assessing the behavior of robust estimates of location in small samples: introduction to configural polysampling," Tech. Rep. No. 185, Series 2, Dept. of Statistics, Princeton Univ., Princeton, NJ.

Stone, C. J. (1975), "Adaptive maximum-likelihood estimates of a location parameter," *Annals Statist., 3,* 267-284.

Tukey, J. W. (1960), "A survey on sampling from contaminated distributions," in *Contributions to Probability and Statistics,* edited by I. Olkin, Stanford Univ. Press, Stanford, CA.

Tukey, J. W. (1976), "Useable resistant/robust techniques of analysis," *Proceedings of First ERDA Statistical Symposium,* Los Alamos, New Mexico, Nov. 3-5, 1975, edited by W. L. Nicholson and J. L. Harris, Battelle Pacific N.W. Labs.

Wegman, E. J. and Carroll, R. J. (1977), "A Monte Carlo study of robust estimators of location," *Comm. Statist.-Theor. Meth., A6,* 9, 795-812.

Whittle, P. (1953), "Estimation and information in stationary time series," *Arch. Math., 2,* 423-434.

Whittle, P. (1962), "Gaussian estimation in stationary time series," *Bull. Int. Stat.*, *39*, 105-129.

Yohai, V. J. and Bustos, O. (1982), "Robust RA-type estimates for ARMA models," Tech. Rep., Dept. of Mathematics, Ciudad Universitaria, Buenos Aires, Argentina.

END

FILMED

2-85

DTIC